— IRONHACK

A/B Testing and Optimization Strategies

Alfons Marques
Lead Teacher @IronHack

AGENDA

1. INTRODUCTION
2. AB TESTING & OPTIMIZATION FUNDAMENTALS
3. AI'S IMPACT ON RETAIL
4. Q&A

INTRODUCTION

Introduction to Retail Evolution

Catalog Shopping Emerges

(Late 1800s)

Launch of mail-order catalogs by Montgomery Ward (1872), retail expands beyond stores.

Barcodes & electronic transactions (EDI)

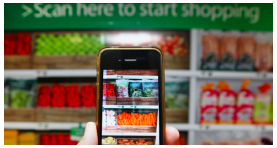
(1970s)

introduction of barcodes and e EDI in the 1970s, revolutionizing inventory management and checkout processes.

Amazon personalized recommendations

(2000s)

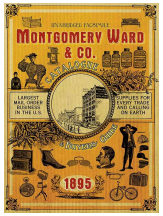
Amazon disrupts the traditional retails with user recommendations based on previous purchases.



The Origins of Retail

(early 1800s)

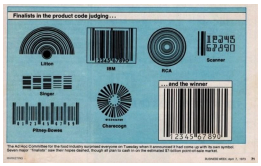
Small family-owned stores, general stores, and the first department stores.



Rising of Shopping Malls

(1950 - 1960)

Post-WWII suburban areas led to the rise of shopping malls, creating a new shopping experience focused on convenience and variety.



E-commerce Begins

(1990s)

The birth of e-commerce with the rise of the internet, and the launch of online retailers like Amazon (1994).

amazon

The Omnichannel Era

(2010s)

Retail seamlessly blends the digital and physical, with AI and AR enhancing every interaction, creating a unified, personalized shopping experience across all channels.

IRON HACK

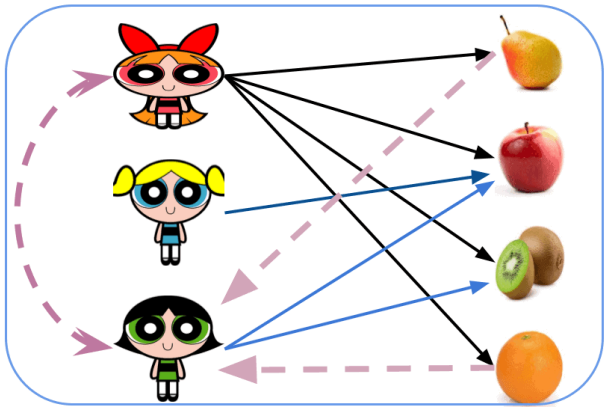
INTRODUCTION

Personalized Recommendations: A Game-Changer

Collaborative filtering →

The goal is to identify patterns and similarities among users. **By finding users with similar tastes and preferences, Amazon can recommend products one user has liked to others who share similar interests.**

Main data points used are **customer's purchase history**, **browsing behavior** and **ratings**.



Amazon uses user-based collaborative filtering to identify users with similar tastes and preferences based on their historical behavior. For instance, if User A and User B have purchased similar books or rated similar movies positively, the system infers that they share similar preferences. Consequently, if User A buys a new book, the system may recommend it to User B based on their past similarities.

In this example, the purchase history shows that the green and red girls have bought the same fruits. In contrast, the green girl shares only one product with the blue girl. So, the system will recommend the green girl other products that the red girl bought.

Amazon also uses content-based filtering. **Content-based filtering analyzes the characteristics of the products, such as their titles, descriptions, categories, and attributes.** By understanding the content of each item, Amazon can recommend similar products based on their features and characteristics.

IRON HACK

INTRODUCTION

The broader impact on retail

Post-Amazon, retailers globally harnessed AI for customer-centric shopping experiences, driving competition in the e-commerce space.

Innovations such as AI-powered chatbots, personalized search results, and predictive analytics became retail standards.

The integration of AR in shopping apps allowed virtual product trials, enhancing decision-making and engagement.

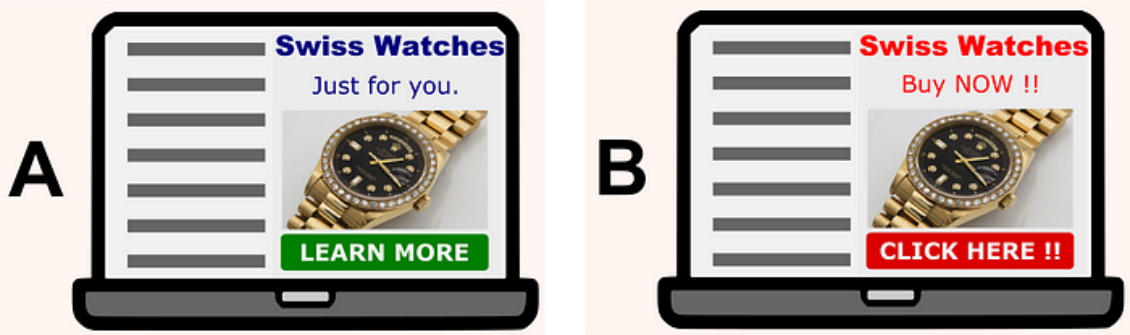
Omnichannel strategies unified online, mobile, and in-store experiences, ensuring consistent and tailored customer interactions.

IRON HACK

AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing

A/B testing, or **split testing**, is a method of comparing two versions of a webpage or app against each other to determine which one performs better. **In classic statistics, we call it hypothesis testing. Optimization** is the ongoing process of using A/B testing results and other data to continuously improve the user experience and achieve business objectives.



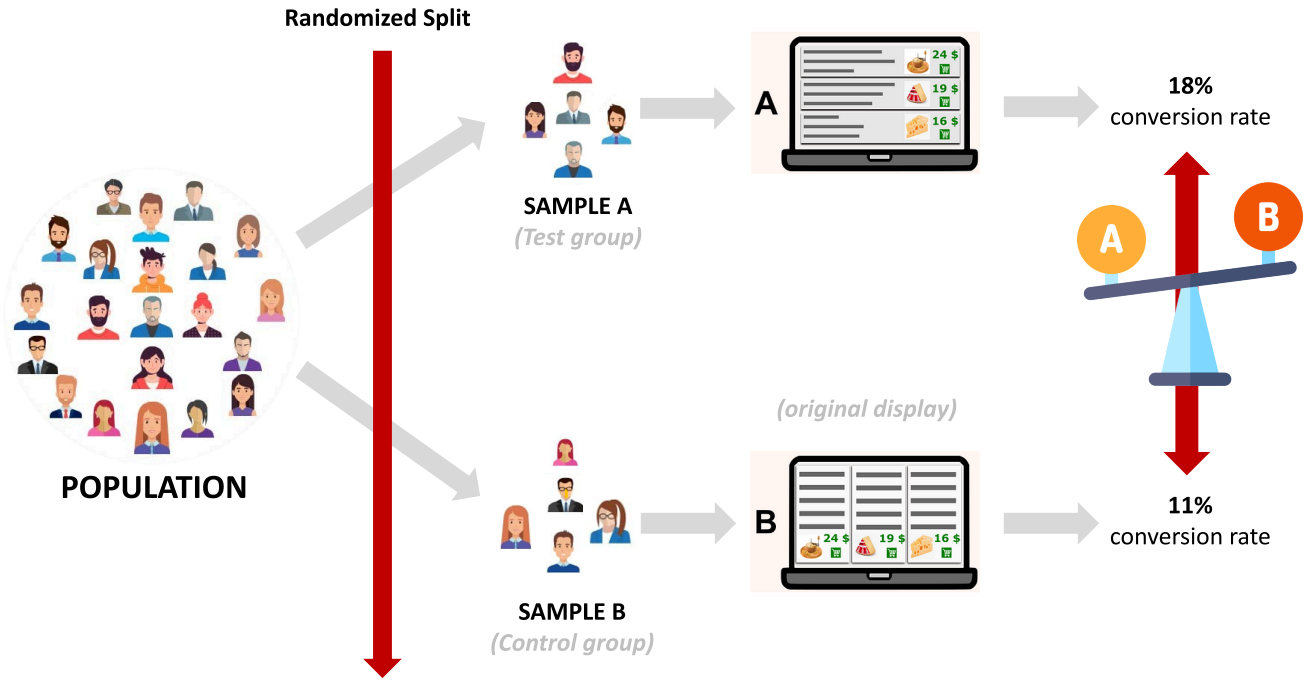
What KPIs do we usually look at here?

- Discrete or binomial metrics (0 or 1 values only possible): *Click-through rate (if a user is shown and advertisement, do they click on it?), conversion rate (if a user is shown an advertisement, do they convert into customers?) and Bounce rate (if a user visits a website, is the following visited page on the same website?)*
- Continuous metrics or non-binomial metrics: *Average revenue per user (how much revenue does a user generate in a month?), Average session duration (for how long does a user stay on a website in a session?) and Average Order value (what is the total value of the order of a user?)*

IRON HACK

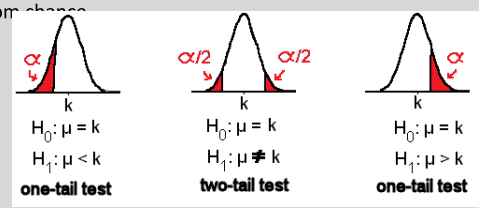
AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing



Minimum Detectable Effect (MDE) is the smallest change or difference you want to be able to spot with your test. It helps decide how many people or items you need to include in your test to confidently say there's a real change and it's not just chance.

With the data we collected from the activity of users of our website, we can compare the efficacy of the two designs A and B. Simply comparing mean values wouldn't be very meaningful, as we would fail to assess the **statistical significance** of our observations. It is indeed fundamental to determine how likely it is that the observed discrepancy between the two samples originates from chance.



Our **null hypothesis H0** is that the two designs A and B have the same efficacy, i.e. that they produce an equivalent click-through rate, or average revenue per user, etc. The statistical significance is then measured by the **p-value**, i.e. the probability of observing a discrepancy between our samples at least as strong as

$$p_{val} = p(\text{data at least as extreme as actual observation} \mid H_0)$$

The p-value is a measure used in statistics to help you determine whether your results are due to chance. If you have a p-value of 5%, it means that if there were no real effect or difference (like no difference between two groups you're comparing), **you would still see the result you got (or something more extreme) about 1 out of 20 times just by random chance.** It's like flipping a coin and getting heads many times in a row – a low p-value suggests it might be a weighted coin rather than a fair one.

AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing

Choosing a statistical test for A/B testing is like picking the right key for a lock, ensuring we correctly interpret our data.

•For categories (like survey responses):

- Use **Pearson's chi-squared test** when you have a lot of data.
- Use **Fisher's exact** test for smaller datasets.

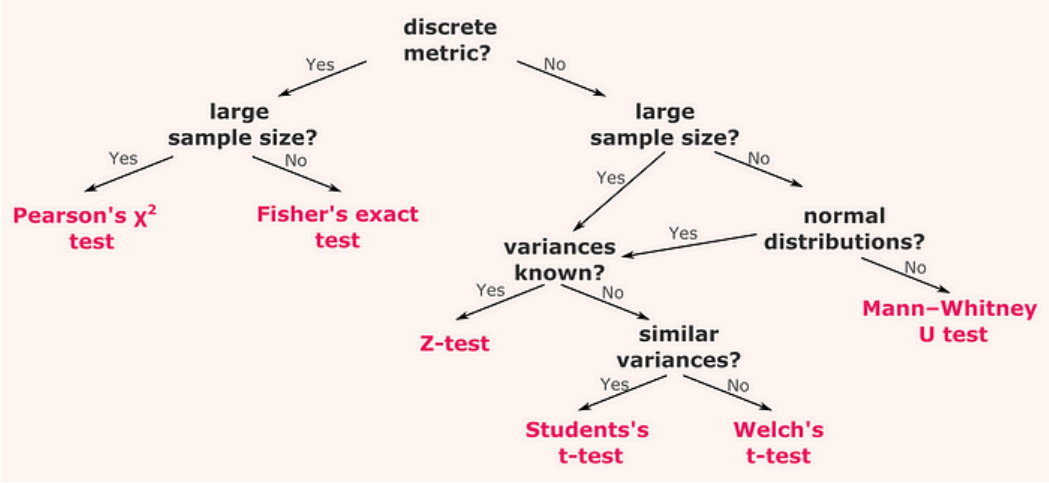
•For numbers (like test scores):

- If you know how spread out the data is (variance), consider a **Z-test**.
- If you're unsure about the spread, but it's roughly equal between groups, use a **Student's t-test**.
- If the spread is unequal, pick **Welch's t-test**.

•If the data is skewed or unusual:

- The **Mann-Whitney U test** works well when data doesn't follow a bell curve.

This guide helps ensure we choose a test that fits our data's story, giving us confidence in our A/B testing results.



AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing

jupyter

IRON HACK

AI'S IMPACT ON RETAIL

Making A/B Testing more advanced

This slide presents a brief overview of three main types of machine learning tasks: classification, regression, and clustering.

Classification	To which category does this data point belong?	Medical diagnosis: does this tissue show signs of disease? Banking: is this transaction fraudulent? Computer vision: what type of object is in this picture? Is it a person? Is it a building?
Regression	Given this input from a dataset, what is the likely value of a particular quantity?	Finance: what is the value of this stock going to be tomorrow? Housing: what would the price of this house be if it were sold today? Food quality: how many days before this strawberry is ripe? Image processing: how old is the person in this photo?
Clustering	Which data points are similar to each other?	E-commerce: which customers are exhibiting similar behaviour to each other, how do they group together? Video Streaming: what are the different types of video genres in our catalogue, and which videos are in the same genre?

IRON HACK


Q&A TIME
imgflip.com